1. Define Perplexity and Probabilities

Let's assume we have M models with perplexities $P_1, P_2, \ldots, P_M$ and their respective predicted probabilities for a word w_i are $p_1, p_2, \ldots, p_M$.

The perplexity of model k is given by:

$$P_k = 2^{H_k}$$

where H_k is the cross-entropy of model k and is related to the probability p_k assigned to the correct word w_i as follows:

$$H_k = -\log_2 p_k$$

Thus, perplexity P_k is:

$$P_k = 2^{-\log_2 p_k} = \frac{1}{p_k}$$

2. Combine Predictions

The combined probability for the correct word w_i when averaging the predictions from M models is:

$$p_{\text{combined}} = \frac{1}{M} \sum_{k=1}^{M} p_k$$

3. Calculate Combined Perplexity

The cross-entropy H_{combined} for the combined model is:

$$H_{\text{combined}} = -\log_2 p_{\text{combined}}$$

and the perplexity P_{combined} of the combined model is:

$$P_{\text{combined}} = 2^{H_{\text{combined}}} = 2^{-\log_2 p_{\text{combined}}} = \frac{1}{p_{\text{combined}}}$$

4. Relate to Harmonic Mean

To relate this to the harmonic mean of individual perplexities, we need to consider that:

$$p_k = \frac{1}{P_k}$$

Thus, substituting p_k into the formula for the combined probability:

$$p_{\text{combined}} = \frac{1}{M} \sum_{k=1}^{M} \frac{1}{P_k}$$

So, the combined perplexity P_{combined} becomes:

$$P_{\text{combined}} = \frac{1}{p_{\text{combined}}} = \frac{M}{\sum_{k=1}^{M} \frac{1}{P_k}}$$

This is the **harmonic mean** of the individual perplexities:

$$P_{\text{combined}} = \frac{M}{\frac{1}{P_1} + \frac{1}{P_2} + \cdots + \frac{1}{P_M}}$$

Summary

To summarize:

- **Individual Perplexities**: $P_k = \frac{1}{p_k}$, where p_k is the probability assigned by model k.
- **Combined Probability**: $p_{\text{combined}} = \frac{1}{M} \sum_{k=1}^{M} p_k$.
- **Combined Perplexity**: $P_{\text{combined}} = \frac{1}{p_{\text{combined}}}$, which simplifies to the harmonic mean of the individual perplexities.

Therefore, **the overall perplexity of multiple models is the harmonic mean of the individual perplexities**. This is because perplexity is inversely related to the geometric mean of predicted probabilities, and combining models involves averaging probabilities, which aligns with the harmonic mean approach.